HUMAN FACTORS CONSIDERATIONS
IN THE ASSESSMENT OF
NONDESTRUCTIVE EVALUATION (NDE) RELIABILITY

Ward D. Rummel

Martin Marietta Aerospace
Denver Aerospace

INTRODUCTION

The overall performance level for an NDE operation is dependent on the NDE material, equipment, processes (methodology) and human skills applied to the operation. It is important to understand and consider human factors elements and contributions to NDE applications in the improvement of applications, in the design and validation of new applications, in automating portions of task performance, and in the development of modeling tools for the prediction of task performance for existing and new applications.

Although human factors variables have been cited in various NDE capabilities studies, the human factors contribution has not often been separated and rigorously addressed as a separate issue (due, in part, to the difficulties in isolating the human contribution from other NDE operations variables). Some classical work relating human factors to functional performance may be cited as a basis for exploring human influences on NDE performance (1,2,3 and 4). This paper explores consideration of some classical methods for potential application to the characterization of human factors in NDE performance.

THE MAKING OF AN NDE OPERATOR

For purposes of this discussion, let us describe optimum or ideal NDE performance as that capability of a proven NDE process to detect an anomaly when carried out by an expert operator (a qualified, trained, skilled and experienced operator). An industrial worker is said to be skilled when he is qualified to carry out a trade or craftwork involving knowledge, judgement and manual deftness, usually acquired as a result of long training, whereas an unskilled worker is not expected to do anything that cannot be learned in a relatively short period of time. The psychological use of the term is wider, and is concerned with factors which go to make up competent, expert, rapid and accurate performance (3).

SKILLS TRAINING AND DEVELOPMENT

After the basic qualifications for entry into a skilled classification are met, the formal training process is initiated. Classroom instruction is used to impart specific knowledge of the principles, limitations and applications of the skill. The instruction is primarily procedural rather than educational in content. The content of the training is generally limited to a narrow area specifically related to the required application of the skill at the user facility. Classroom training is usually followed by an examination to ascertain the level of near-term understanding and recall of the specific facts and procedures of interest. Classroom training may be followed by laboratory training to develop necessary motor and sensory skills under closely supervised conditions. An examination covering the level and quality of performance may be administered. Skill development is completed by a period of "on the job training" during which performance is closely monitored and verified by a skilled supervisor or co-worker. After a further period of experience and practice, the knowledge and skill information are "compiled" in the brain of the individual operator and a level of expertise is established. Such expertise becomes part of the "long term memory" of the operator and operations involving application of the "compiled" expertise can be carried out with less conscious attention to the coordinated elements of the task.

HUMAN ERRORS IN NDE OPERATIONS

Errors in performance by skilled operators may be classified as: SYSTEMATIC ERROR (consistent offset from ideal performance); ERRORS IN PRECISION (consistent, but random, variations in performance about a norm); SPORADIC ERRORS (an occasional occurrence varying significantly from the norm). A systematic error is experienced as a constant offset, for example, a wind condition when target shooting. An error in precision is experienced as the scatter pattern when target shooting. A sporadic error is typified by landing a jet aircraft in a lake, 1000 feet short of the runway (5).

An NDE task is an exercise in conditional probability with two error modes: TYPE I, failure to find a flaw when a flaw is present (False Negative) and TYPE II, finding a flaw when no flaw is present (False Positive). The result of all types of classified errors are manifested as Type I or Type II errors in the overall results from an inspection task.

INFLUENCE OF HUMAN FACTORS ON THE PROBABILITY OF DETECTION

A recognized method characterizing the performance capability of an NDE method is by quantification and plotting outcome as a probability of detection (POD) curve as shown in Figure 1. An ideal NDE process yields a POD curve with a constant, high level of detection at large flaw sizes and a sharp drop off in discrimination at small flaw sizes. Sporadic errors in detection at large flaw sizes will result in data scatter at large flaw sizes and is usually due to sporadic human error (Figure 2). Such errors are often associated with drowsiness, lack of interest, lack of motivation, fatigue, boredom, monotony or state of arousal (vigilance). Such errors may be minimized by attention to the factors responsible for the error occurrence and by redundant inspections (6).

Errors in precision are indicated by data scatter at the transition region of the POD curve and can be caused by slight variations in processing, by inexperience of the operator or by a shift in decision criteria (usually due to a lack of confidence) during the processing operation. Experience, expert skill development and well defined and recognized decision criteria will minimize this mode of human error. (See Figure 3.)

Systematic errors are indicated by a shift in the threshold transition point on the POD curve when inspection is carried out on identical components by two different operators or by a single operator operating with two different sets of decision criteria. (See Figure 4.) Differences in performance may be due to a difference

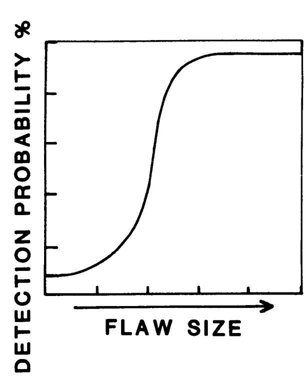

Figure 1. Idealized POD Curve.

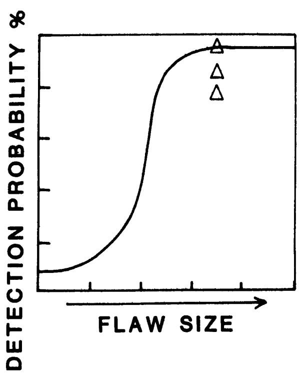

Figure 2. Variable Detection (Sporadic Errors).

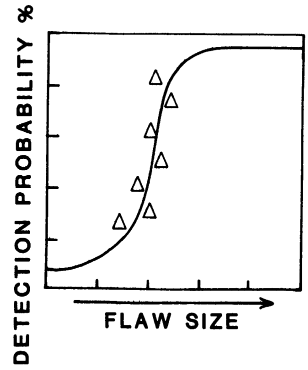

Figure 3. Impact of Errors of Precision on POD.

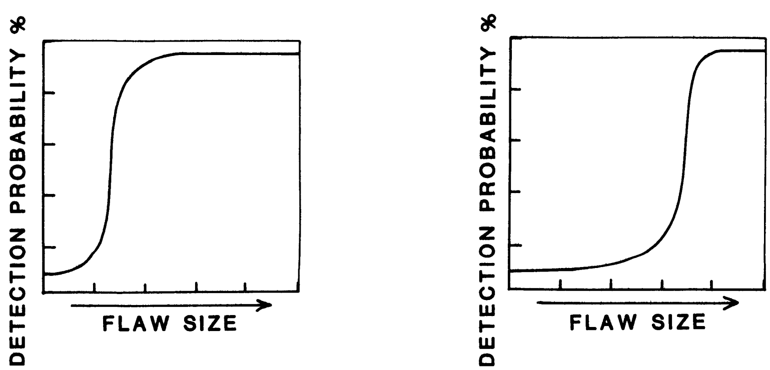

Figure 4. Two POD Curves--Different Decision Criteria.

in skill and/or decision criteria input by the operators; or may be due to differences in processing materials, processing equipment, calibration standards or procedures for the two inspections. Proper training and direction regarding decision criteria have been shown to reduce systematic errors between inspections and between operators (7,8, and 9).

When two such data sets are combined, the resultant POD curve is as shown in Figure 5. Figure 5 has no sharp threshold transition point and is difficult to interpret due to the multiple variations that affect the POD as a function of flaw size. Such curves are frequently associated with round robin test programs (10 and 11).

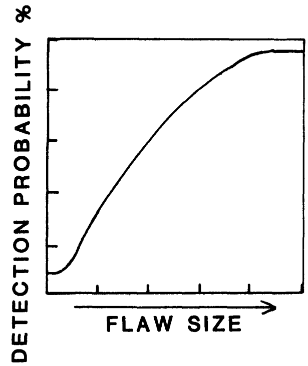

Figure 5. POD for Combined Results--Different Decision Criteria.

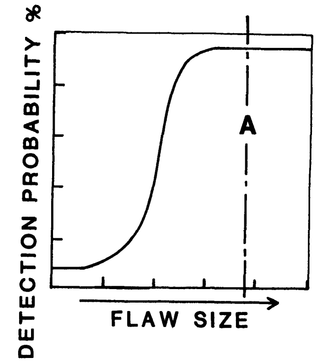

Figure 6. POD for High Detectability (Point A).

SIGNAL DETECTION AND CHARACTERIZATION

Quantification of signals from known standards is suggested as a method of approaching the problems of variation in performance between facilities due to calibration differences, equipment differences, etc., as well as that of predicting a level of ideal performance.

If we refer to the POD curve as shown in Figure 6, it is obvious that large flaws are readily detected by the NDE techniques with a "high signal threshold". The point of interest is that near the threshold detection point where it is necessary to detect faint, infrequent signals from background noise. Such a problem is similar to that addressed in characterizing operator performance with radar equipment. The requirement to detect faint, infrequent signals was addressed by studies of human variations in signal detection (ie. continued performance as a function of time) and variations in the decision process (ie. variations in performance as a function of signal to noise ratio.

SIGNAL DETECTION AND DECISION THEORY

The work of primary interest in NDE engineering and in automation of NDE processes is that of signal detection and analyses. Early application of statistical decision theory to signal detection was made by Tanner and his associates (3, 7) in which they proposed that signals have to be detected against a certain amount of background noise and that the signal required to secure any degree of accuracy will increase with the noise level. The discrimination of a process can be specified in terms of signal-to-noise ratio.

If we consider a single operating point near the threshold discrimination level of a probability of detection (POD) curve, as noted in Figure 6, noise can be successfully discriminated from signal plus noise and inspection success can be predicted. If such an inspection is repeated successively, a probability density function versus signal amplitude may be generated as shown in Figure 7. It is clear that the noise distribution is separated from the signal plus noise distribution and positive discrimination between the two can be made by choosing a decision criterion level shown as level A. The actual success of the discrimination is dependent on the detection criteria applied to the observed signals. If the decision criteria (signal required to call a positive detection) is set too high (moved to the right as in level B), successful discrimination will result most of the time, but some flaws will be accepted with resultant Type I errors. If the decision point is set too low (moved to the left as in level C), successful detection will result all of the time, but some noise signals will be incorrectly identified as flaws and good parts will be rejected (Type II errors). For a given nondestructive evaluation method, operating point and flaw size, the probability of detection (true positive) and the probability of a false alarm (false positive) may be plotted as shown in Figure 8. The diagonal line is that data form for random chance with no discrimination. The upper left coordinates describe the performance of a perfect inspection with total discrimination. This method of plotting was originated and termed the "Relative Operating Characteristic or ROC" curve by Tanner and Swets (4). Points along the ROC curve constitute a characteristic inspection state at varying decision criterion values. For a given operating point, quantitative values for detection success and false call rates can be obtained. It is clear, from such analysis, that the human operator can be very reliable, if the signal and noise acceptance criteria are established at the proper level and if a high degree of discrimination is attained by the specific NDE method. Conversely, the boundary conditions of the method and decision criteria, provide primary physical limits to the capability of a given inspection.

SUMMARY AND CONCLUSIONS

The human operator is usually reliable in performing a wide variety of tasks but must operate within the boundary conditions of his capabilities and within the boundary conditions set by the physical limits of the task to be performed. NDE tasks are complex and require knowledge, skill, experience and dexterity for optimum performance. When detection and discrimination are not attained by an NDE process, the most frequent cause stated is that of operator errors, which can and do occur. Proper attention must be given

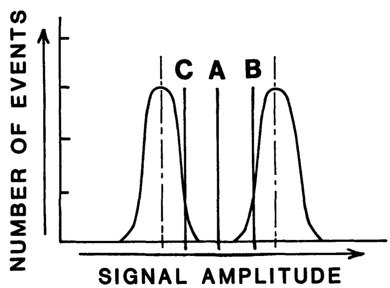

Figure 7. High Discrimination Level-Point A.

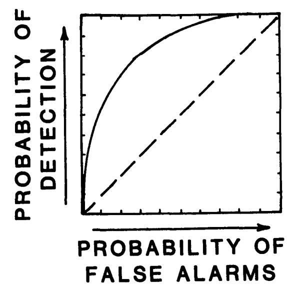

Figure 8. ROC Curve-Single Point on POD Curve.

to those NDE processes involving human operators to assure that the tools, working conditions and environment are commensurate with the task being performed.

The most frequent cause for unreliable NDE performance that has been observed by the author is that of improper NDE engineering. In many cases, the NDE method selected is incorrect or was not qualified and controlled to the level necessary to obtain the required discrimination. In other cases, the NDE equipment, materials and processes are not controlled and do not provide discrimination to the level qualified. In short, human factors in NDE involves NDE engineers, materials and process control engineers, calibration and maintenance personnel, training personnel, supervision--and the performing operator. We have no right to expect reliable detection and discrimination by the human operator (or by an automated unit), when an NDE process has been altered up-stream of a detection opportunity. NDE reliability involves the entire NDE team and that team may be much larger than you think. The science and technology being developed and reported in this symposium must be incorporated as tools by the NDE team in the form of improved requirements specifications and in improved understanding of the capabilities and limitations of NDE for reliable detection and discrimination.

REFERENCES

1. W. P. Tanner and J. A. Swets, A Decision-making Theory of Visual Detection, Psychology Review 61:404-409 (1954).

2. J. A. Swets, W. P. Tanner, and T. G. Birdsall, Decision Processes in Perception, Psychology Review 68:301-340 (1961).
3. A. T. Welford, "Fundamentals of Skill," Methuen & Co. Ltd., London (1968).
4. J. A. Swets, ed., "Signal Detection and Recognition by Human Observers," Wiley, New York (1964).
5. S. Deutsch, Human Factors and Measurement Uncertainties, ASME Transactions - Nondestructive Evaluation (NDE) and Human Factors, October 12-13, 1981.
6. J. N. Yang and R. C. Donath, Improving NDE Capability Through Multiple Inspections with Application to Gas Turbine Engine Disks, Air Force Wright Aeronautical Laboratories Report, AFWAL TR-82-411, June 1983.
7. E. L. Caustin, Nondestructive Testing Demonstration, North American Rockwell, Contract No. F33657-70-C-0800, 1971.
8. H. R. Chin Quan and I. G. Scott, Operator Effects in NDI--Part I, Eddy Current Measurements on Aluminum Alloy Blocks Containing Edge Slots, Metallurgy Note 82, Aeronautical Research Laboratories, Australian Defense Science Service, 1972.
9. J. C. Herr and G. L. Marsh, NDT Reliability and Human Factors, Materials Evaluation 36:13:41-46 (December 1978).
10. J. T. Fong, ed., "Nondestructive Evaluation (NDE): Reliability and Human Factors," ASME Transactions, October 12-13, 1981.
11. J. T. Fong, S. J. Bush, and O F. Hedden, eds., "NDE Reliability Through Round Robin Testing," ASME Transactions, June 19-21, 1983.

DISCUSSION

From the Floor: I've been involved in NDE certification and this is relevant to problems that we are having in NDE certification. My experience of almost 30 years of NDE is that we have got very few poor NDE operators in the industry. We have had a lot of problems involving NDE operators. The big problem has been in management control, lack of appropriate training provided to operators, and a lack of materials capability, not the NDE portions. It's harder to relate the NDE capability to materials properties. In other words, to anticipate what the defect is and what all defects would be in that particular problem that we're working with. We should be able to, from our mind's eye, anticipate what will happen from a problem presenting itself and then being able to read from the data what that problem actually is. In other words, we anticipate.

It is easy if all the parameters are known and one knows exactly where to look. But if you don't know where to look, sometimes the flaws don't show up. So what happens is we have NDE operators being blamed because they are not qualified and lack certification. Managers expect to get all their NDE problems taken care of by certification, but it is also necessary to do a better job of planning the NDE training for them in the materials they want to inspect and in the nature of the test they are going to be asked to implement. If these changes were adopted, we will have a lot less trouble.

W.D. Rummel: I would like to add to that. We do a pretty good job in the front-end classroom training but in the on-the-job training and in the feed-back, sometimes it is a little unclear what's required to actually do the job.

In one particular case in which we were doing a reliability study at the facility, the operators were absolutely delighted to participate in this exercise because this is the first time that they had been able to operate on parts with no flaws. They viewed it as feedback and part of their training.

I.G. Scott (Department of Defence, Australia): Would you care to comment further on rewards and penalties? That's something that's pretty difficult to build into at NDE.

W.D. Rummel: It is very difficult to build into the human operations also because the reward and penalty cannot be isolated to that one quadrant where we are talking about people. We must make sure that all factors are in. If we make a conscious effort to make sure all factors are doing that, then rather than rewards and penalty, we remove those operators who indeed can't perform and we will probably have to pay those operators who are performing more and that's one of the suggestions that's been made. In order to avoid turnover in an area, you have to pay those operators more and maintain their skill.

I.G. Scott: Can I qualify that question a little? You have answered it perfectly. In abbreviating the question, I perhaps didn't put it as clearly as possible. The penalty for a mistake on Columbia is far greater than the penalty for a mistake on your own car. Would you care to comment on that?

W.D. Rummel: That's the risk analysis part of it. Risk analysis is viewed differently in different aspects. Unfortunately, we too often identify with the liability that goes with the consumer product as opposed to the more spectacular kind of a problem that we might have if Columbia falls through or is not done properly. When it's taken out of the hands of those who are knowledgeable about the performance of the hardware and relegated to the lawyers, then we have an entirely different problem. So I think that's the difference. I think we are talking apples and oranges in the two risks.

B.W. Staff (Lockheed-Georgia Co.): There's been a number of studies and a fair amount of work in the last few years on probabilistic studies for NDE as well as getting into fracture mechanics, but the only suggested way of handling the false call is associated with the real cost associated with the real application with that false call, should it result in a major catastrophe. Is there any other way that you know of that this can be handled?

W.D. Rummel: I think the proper place for the false calls is in the establishment of the technique on the front end or defining what that relative operating characteristic curve is or the decision point and designing both the training and the actual application around that. Then you must make a conscious decision to accept the false call rate that goes with that reliability.

B.W. Staff: But that involves the lawyer and the manager up front.

W.D. Rummel: In the case of a false call, if you make that decision up front without the lawyer and manager, what it means is that you may do a redundant inspection or you may do a complementary inspection to alleviate any problems that may result from that false call.

D.E.W. Stone (Royal Aircraft Establishment): How much do you think that management can do to alleviate the aspects of boredom? Perhaps by introducing the occasional deliberate fault? In many situations where they have been looking for a defect for several years and never found one, I think they'll never find one with that particular technique. Can't management improve on this?

W.D. Rummel: Well, management can improve on the front end by making sure we have the proper technique that's capable of discovering. As I mentioned, we can have redundancy and then we can work the other part of the human factors problem to make sure that the job assignment is such that we minimize the boredom, we have some proper rewards and penalties, so to speak, or rewards, attention given when a flag is actually raised.

J.C. Coffey (Central Electricity Generating Board, England): You wouldn't go so far as to slip in a deliberate defect?

W.D. Rummel: Any time we try to game the operator (there has been some of that done), we find that we aren't as smart as the operators are on the floor.